~~FILE COPY~~

RETURN TO Stat. Eng. LAB
401 South
N.B.S.

NBS PROJECT **NBS REPORT**

~~Section 11.3~~

~~FILE COPY~~

~~Do not remove~~

1103-21-5118

3950

18 February 1955

Preliminary Report

on

COMPUTING THE ANALYSIS OF VARIANCE OF FACTORIAL EXPERIMENTS ON AUTOMATIC COMPUTERS

by

J. M. Cameron
Statistical Engineering Laboratory

to

Chemical Corps Biological Laboratories
Camp Detrick, Maryland

COMPUTING THE ANALYSIS OF VARIANCE OF FACTORIAL EXPERIMENTS ON AUTOMATIC COMPUTERS

(A Preliminary Report)

by

J. M. Cameron
Statistical Engineering Laboratory
National Bureau of Standards

The purpose of this note is to describe the method of Yates* which seems well suited for computing the analysis of variance of factorial experiments on large scale computers. The analysis for the 2^n and 3^n series is given and the analysis for the general factorial indicated. The detailed analysis of examples of the 2^n and 3^n designs are given.

1. Analysis of variance of the 2^n factorial designs.

In a factorial design the effects of a number of factors are investigated simultaneously. In the 2^n factorial design there are n factors (such as temperature, dilution, etc.) each of which is studied at two levels (e.g. 20° and 30° for temperature). Test conditions are set up corresponding to each of the possible combinations of the two levels of the n factors (i.e. 2^n combinations in all) and an observation is recorded for each.

The 2^n combinations can be designated by a series of n subscripts which are either 0 or 1 depending on whether the factor is at its low or high level. For example for $n = 3$ as shown below each of the $2^3 = 8$ possible combinations can be designated by X_{abc}, where a, b, and c are either 0 or 1.

In the procedure given here it is necessary that the observations be presented in a column of 2^n values in the order shown below for the case $n = 3$. The extension to other values of n is obvious.

---

* F. Yates, "Design and Analysis of Factorial Experiments", Imperial Bureau of Soil Science, Tech. Comm. No. 35, Harpenden, 1937.

Designation of observation	Level of factor A	B	C
X_{000}	0	0	0
X_{100}	1	0	0
X_{010}	0	1	0
X_{110}	1	1	0
X_{001}	0	0	1
X_{101}	1	0	1
X_{011}	0	1	1
X_{111}	1	1	1

Once this column of data is in the machine a column of "sums and differences" also containing 2^n values is obtained from it by tabulating the 2^{n-1} sums of successive pairs of observations followed by the 2^{n-1} differences between the elements of the same pairs. Thus for $n = 3$ we have

Observations	1st "Sums and Differences"
X_{000}	$X_{000} + X_{100}$
X_{100}	$X_{010} + X_{110}$
X_{010}	$X_{001} + X_{101}$
X_{110}	$X_{011} + X_{111}$
X_{001}	$X_{000} - X_{100}$
X_{101}	$X_{010} - X_{110}$
X_{011}	$X_{001} - X_{101}$
X_{111}	$X_{011} - X_{111}$

The same process is repeated on the first column of "sums and differences" to form a second column of "sums and differences", and so on for each successive column so formed until the n-th column of "sums and differences" is obtained.

The square of an entry in the n-th column of "sums and differences" divided by 2^n corresponds to a single degree of freedom in the analysis of variance. The single degrees of freedom for the several main effects and interactions come out in the following sequence

CF (correction factor for the mean)
A
B
AB
C
AC
BC
ABC
D
AD
BD
ABD
CD
ACD
BCD
ABCD

E
AE
BE
ABE
CE
ACE
BCE
ABCE
.
.
.
(etc.)

The entries in the n-th column of "sums and differences" divided by 2^{n-1} gives an estimate of the average difference between the levels of a factor.

Computational checks.

1. The sum of the entries in the n-th column of "sums differences" is equal to $2^n \, x_{00000...0}$. (i.e. 2^n times the leading element in the data as presented to the machine)
2. The sum of the squares of the entries in the n-th column is equal to 2^n times the sum of squares of the elements of the original column of observations.

Analysis of Variance for 2^n Factorial Design: Example for $n = 4$

Level of factor A	B	C	D	observed value* for treatment combination
0	0	0	0	60
1	0	0	0	20
0	1	0	0	83
1	1	0	0	59
0	0	1	0	19
1	0	1	0	77
0	1	1	0	13
1	1	1	0	39
0	0	0	1	5
1	0	0	1	26
0	1	0	1	27
1	1	0	1	85
0	0	1	1	25
1	0	1	1	47
0	1	1	1	86
1	1	1	1	76

Input

---

*Taken from table of random numbers.

PRINT OUT

X_{0000}

[illegible] ABCD	Observed value X	Step 1	Step 2	Step 3	D = Step 4	D^2	$\frac{D}{2^{n-1}}$ [illegible] diff between levels of a factor		$\frac{D^2}{2^n}$ analysis of variance
							* correction		
0000	60	80	232	370	747	558,009	93375	for mean	34875.5625
1000	20	192	108	377	-111	12,321	-13.875	A	770.0625
0100	83	96	193	-20	-189	35,721	-23.625	B	2 232.5625
1100	59	52	234	-91	-11	121	-1.375	AB	7.5625
0010	19	31	64	-18	-17	289	-2.125	C	18.0625
1010	77	112	-94	-131	81	6,561	10.125	AC	410.0625
0110	13	72	-79	-16	-97	9,409	-12.125	BC	588.0625
1110	39	162	-12	5	117	13,689	14.625	ABC	855.5625
0001	5	-40	-62	74	-7	49	-0.875	D	3.0625
1001	26	29	-49	-91	71	5,041	8.875	AD	315.0625
0101	27	-58	-81	149	153	23,409	19.125	BD	1463.0625
1101	35	-26	-90	-67	-21	441	-2.625	ABD	27.5625
0011	25	-21	16	-106	165	27225	20.625	CD	1701.5625
1011	47	-58	-82	9	215	46225	26.875	ACD	2889.0625
0111	86	-22	37	49	-115	13,225	-14.375	BCD	826.5625
1111	76	10	-32	69	-21	441	-2.625	ABCD	27.5625
Sum	747				960	752,176	120.000	Total	47,011.000
Sum of Squares	47,011				752,176				

Checks: ① $\Sigma D = 2^n X_{0000}$ $960 = 2^4(60)$

② $\frac{\Sigma D^2}{2^n} = \Sigma X^2$ $752,176 = 2^4(47,001)$

* note that this value is Twice the grand average

2. Analysis of variance of the 3^n factorial designs

Let the observations be presented according to the scheme shown here for n = 4.

Designation of observation	Level of factor A	B	C	D
X_{0000}	0	0	0	0
X_{1000}	1	0	0	0
X_{2000}	2	0	0	0
X_{0100}	0	1	0	0
X_{1100}	1	1	0	0
X_{2100}	2	1	0	0
X_{0200}	0	2	0	0
X_{1200}	1	2	0	0
X_{2200}	2	2	0	0
X_{0010}	0	0	1	0
X_{1010}	1	0	1	0
X_{2010}	2	0	1	0
.				
.				
.				
.				
X_{0222}	0	2	2	2
X_{1222}	1	2	2	2
X_{2222}	2	2	2	2

The 3^n observations give rise to 3^{n-1} successive sets of 3 values. A column of "sums and differences" having 3^n elements is formed as follows (1) the sums of the 3 elements of the 3^{n-1} sets are tabulated in order, (2) these are followed by the 3^{n-1} differences between the first and third elements of the sets, and finally (3) the sum of the first and third minus twice the middle value for each of the 3^{n-1} sets are tabulated to complete the first column of "sums and differences".

This same procedure is repeated on the successive columns of "sums and differences" until the n-th column of "sums and differences" is obtained. The square of the elements, D, of this n-th column divided by a corresponding factor, d, gives a single degree of freedom for the main effects or interactions in the analysis of variance table. These single degrees of freedom come out in the following order. (The subscript 1 refers to the linear component and the subscript 2 refers to the quadratic component.)

CF
A_1
A_2
B_1
A_1B_1
A_2B_1
B_2
A_1B_2
A_2B_2
C_1
A_1C_1
A_2C_1
B_1C_1
$A_1B_1C_1$
$A_2B_1C_1$
B_2C_1
$A_1B_2C_1$
$A_2B_2C_1$
C_2
A_1C_2
A_2C_2
B_1C_2
.
.
.
(etc.)

The appropriate divisors for the squares of the entries in the n-th column are given by raising the triple (3, 2, 6) to the n-th power according to the following rule:

for n = 2 $(3, 2, 6)^2 = (3,2,6)\ (3,2,6)$

$$\begin{array}{ccccc} 3 & 2 & 6 & & \\ 3 & 2 & 6 & & \\ \hline 9 & 6 & 18 & & \\ & 6 & 4 & 12 & \\ & & 18 & 12 & 36 \end{array}$$

The sequence of divisors for the corresponding elements of the n-th column being

9, 6, 18, 6, 4, 12, 18, 12, 36

for n = 3 $(3,2,6)^3 = (3,2,6)^2\ (3,2,6)$

```
  9   6  18   6   4  12  18  12  36
  3   2   6
-------------------------------------
 27  18  54  18  12  36  54  36 108
     18  12  36  12   8  24  36  24  72
         54  36 108  36  24  72 108  72 216
```

The sequence of divisors being

27, 18, 54, 18, 12, ..., 72, 108, 72, 216.

The extension to larger values of n is carried on in the same manner. The case n = 4 is given in the worked out example.

Computational checks

(1) The sum of squares of the original observations is equal to the sum $\Sigma \frac{D^2}{d}$.

(2) The sum of the n-th column of "sums and differences" can be checked using the following procedure: from the successive sets of three values of the observation column form a column of the 3^{n-1} quantities obtained by taking 3 times the first element of the set minus the middle element plus the third element. Repeat this process on the column so formed. After n repetitions one final number remains. This number is the check sum for the n-th column of sums and differences.

Combining the individual degrees of freedom

The analysis of variance table is usually written in the form shown here for n = 4.

Analysis of variance table

Factor	Sum of Squares is Sum of	Degrees of Freedom d.f.	mean square
CF	CF	1	
A	A_1+A_2	2	
B	B_1+B_2	2	
AB	$A_1B_1+A_2B_1+A_1B_2+A_2B_2$	4	
C	C_1+C_2	2	
AC	.		$\frac{\text{sum of squares}}{\text{d.f.}}$
BC	.		
ABC	.		
D	.		
AD	etc.	.	
BD		.	
ABD		.	
CD		.	
ACD		8	
BCD		8	
ABCD		16	

In order to convert the column of values $\frac{D^2}{d}$ (corresponding to the 81 individual degrees of freedom for n = 4) into this conventional form for the analysis of variance tables, one can use successive triads of the column of 81 individual d.f. Two columns of values are formed from the 27 triads: (1) the first element of each triad is recorded in sequence in the first column, and (2) the sum of the last two elements of the 27 triads is recorded in the second column.

These two columns are then combined into a single column of 54 elements by writing the second column at the end of the first. This process is repeated n = 4 times and the resulting column is the "sum of squares" column in the standard analysis of variance table.

The degrees of freedom associated with the 16 sums of squares for $n = 4$ may be obtained by raising the couple (1,2) to the 4-th power as follows:

$$(1,2)\ (1,2) = (1,2,2,4)$$

$$(1,2)^3 = (1,2)^2\ (1,2) = (1,2,2,4,2,4,4,8)$$

$$(1,2)^4 = (1,2)^3\ (1,2) = (1,2,2,4,2,4,4,8,2,4,4,8,4,8,8,16)$$

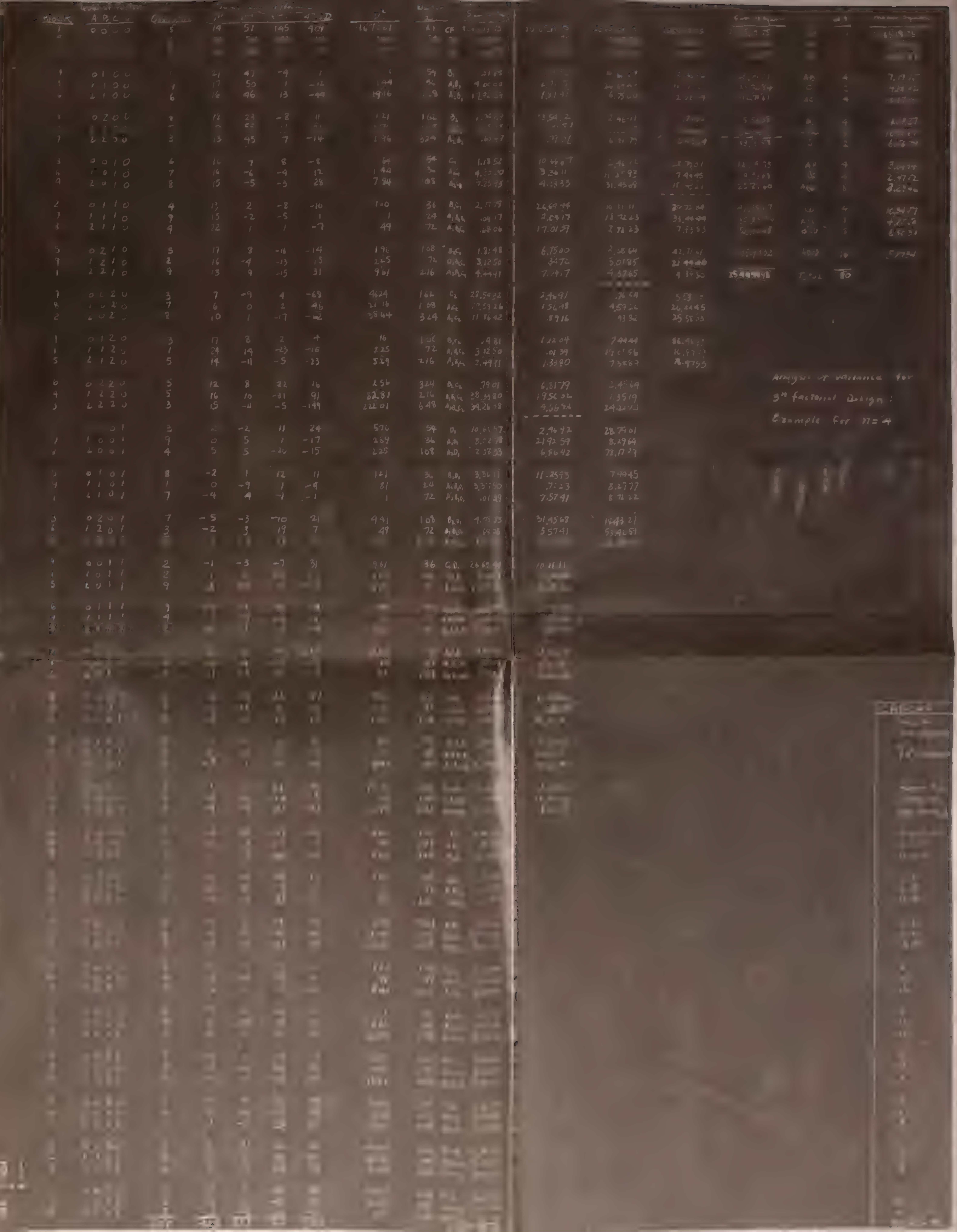
etc.

The mean square is obtained by dividing the "sum of squares" by this divisor.

Binc

1
5
4

4
5



Block	A B C D	Observ.	1st	2nd	3rd	4th D	D²	Divisor		Sum of squares			
1	0 0 0 0	5	19	51	145	409	167281	81	CF	[illegible]	2065.1975	[illegible]	[illegible]
2	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
4	0 1 0 0	1	21	47	-9	1	1	54	B_1	.0185	[illegible]	[illegible]	[illegible]
5	1 1 0 0	4	17	50	1	-12	144	36	A_1B_1	4.0000	[illegible]	[illegible]	[illegible]
6	2 1 0 0	6	16	46	13	-44	1936	108	A_2B_1	17.9259	1.3143	6.7500	[illegible]
3	0 2 0 0	8	18	23	-8	11	121	162	B_2	[illegible]	13.5912	2.9611	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
7	2 2 0 0	3	13	43	7	-14	196	324	A_2B_2	.6047	.7901	[illegible]	[illegible]
3	0 0 1 0	6	16	7	8	-8	64	54	C_1	1.1852	10.6607	2.4612	[illegible]
6	1 0 1 0	7	16	-6	-4	12	144	36	A_1C_1	4.0000	3.3611	11.2593	7.4445
9	2 0 1 0	8	15	-5	-3	28	784	108	A_2C_1	7.2593	4.3333	31.4568	[illegible]
2	0 1 1 0	4	13	2	-8	-10	100	36	B_1C_1	2.7778	26.6944	10.1111	[illegible]
7	1 1 1 0	9	15	-2	-5	1	1	24	$A_1B_1C_1$	.0417	2.0417	18.7223	33.4444
8	2 1 1 0	9	22	1	1	-7	49	72	$A_2B_1C_1$	.6806	17.0139	2.7223	[illegible]
5	0 2 1 0	5	17	8	-16	-14	196	108	B_2C_1	1.8148	6.7500	2.3864	42.7161
9	1 2 1 0	2	16	-4	-13	15	225	72	$A_1B_2C_1$	3.1250	.3472	5.0185	21.4446
1	2 2 1 0	9	13	9	-15	31	961	216	$A_2B_2C_1$	4.4491	7.1917	4.3765	4.3950
7	0 0 2 0	3	7	-9	4	-68	4624	162	C_2	28.5432	2.4691	.7654	[illegible]
8	1 0 2 0	7	6	0	2	46	2116	108	A_1C_2	19.5926	1.5648	4.5926	26.4445
2	2 0 2 0	8	10	1	-17	-62	3844	324	A_2C_2	11.8642	.8916	[illegible]	25.5803
1	0 1 2 0	3	17	8	2	4	16	108	B_1C_2	.1481	1.2204	7.4444	[illegible]
1	1 1 2 0	1	24	19	-23	-15	225	72	$A_1B_1C_2$	3.1250	.0139	[illegible]	[illegible]
5	2 1 2 0	5	14	-11	-5	-23	529	216	$A_2B_1C_2$	2.4491	1.3380	[illegible]	[illegible]
6	0 2 2 0	5	12	8	22	16	256	324	B_2C_2	.7901	6.8179	[illegible]	
4	1 2 2 0	5	16	10	-31	91	8281	216	$A_1B_2C_2$	38.3380	19.5602	1.3519	
3	2 2 2 0	3	15	-11	-5	-149	22201	648	$A_2B_2C_2$	34.2608	[illegible]	[illegible]	
[illegible]	[illegible] 0 1	3	[illegible]	-2	11	24	576	54	D_1	10.6667	2.4692	28.7901	
7	1 0 0 1	9	0	5	1	-17	289	36	A_1D_1	[illegible]	2192.59	8.2964	
[illegible]	2 0 0 1	4	5	5	-20	-15	225	108	A_2D_1	2.0833	6.8642	[illegible]	
5	0 1 0 1	8	-2	1	12	11	121	36	B_1D_1	3.3611	11.2593	7.4445	
4	1 1 0 1	1	0	-9	1	-9	81	24	$A_1B_1D_1$	3.3750	[illegible]	8.2777	
1	2 1 0 1	7	-4	4	-1	-1	1	72	$A_2B_1D_1$	.0139	7.5741	8.7222	
3	0 2 0 1	7	-5	-3	-10	21	441	108	B_2D_1	4.0833	31.4568	[illegible]	
6	1 2 0 1	3	-2	3	19	7	49	72	$A_1B_2D_1$	.6806	5.5741	53.4251	
9	0 0 1 1	2	-1	-3	-7	31	961	36	C_1D_1	[illegible]	10.1111		
1	1 0 1 1	2	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]		
5	2 0 1 1	9	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]		
6	0 1 1 1	9	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]		

Sum of Squares		d.f.	Mean Square
[illegible]	[illegible]	1	[illegible]
[illegible]	AB	4	7.1975
[illegible]	C	2	[illegible]
[illegible]	AC	4	[illegible]
[illegible]	[illegible]	4	[illegible]
[illegible]	[illegible]	2	[illegible]
[illegible]	AD	4	[illegible]
[illegible]	[illegible]	4	2.4712
[illegible]	[illegible]	8	[illegible]
[illegible]	[illegible]	4	[illegible]
[illegible]	[illegible]	8	[illegible]
[illegible]	[illegible]	8	[illegible]
[illegible]	[illegible]	16	[illegible]
[illegible]	Total	80	

Analysis of variance for
3^n factorial Design:
Example for $n=4$

3. Analysis of variance of general factorial design

The proper sequence for presenting the column of observations can probably best be described by giving an example. For a 4 x 3 x 2 factorial the observations are presented in the order corresponding to the following combination of the factors

level of factor		
A	B	C
0	0	0
1	0	0
2	0	0
3	0	0
0	1	0
1	1	0
2	1	0
3	1	0
0	2	0
1	2	0
2	2	0
3	2	0
0	0	1
1	0	1
2	0	1
3	0	1
0	1	1
1	1	1
2	1	1
3	1	1
0	2	1
1	2	1
2	2	1
3	2	1

In general for a k x m x r factorial (k>,m>,r) the observations are put in order so that the k levels of the factor A are recorded in their sequence m r times. The corresponding index for B is obtained by writing each index for B k times and repeating this sequence r times. The C index is obtained by writing each successive index for C k m times. For factorials with other than three factors, the same approach is applied.

In the procedure described here a column of "sums and differences" is formed for the first factor. For a factor at two levels, the column is formed by taking sums and differences in successive pairs as described for the 2^n factorial. For a factor at 3 levels, triads are used and the column of "sums and differences" is formed as described for the 3^n factorials. For factors at 4,5,6, or more levels a "sums and differences" column is formed by operating on sets of 4, 5, 6, etc.

The sums and differences to be used can be obtained from a table of orthogonal polynomials (See D. B. Delury*).

The following table shows the linear functions of the sets of 2, 3, 4, ... corresponding to factors having 2, 3, 4, ... levels.

No. of levels of factor	Coefficients of linear functions	d_{ij}
n = 2	1+1	$2 = d_{21}$
	1-1	$2 = d_{22}$
n = 3	1+1+1	$3 = d_{31}$
	1+0-1	$2 = d_{32}$
	1-2+1	$6 = d_{33}$
n = 4	1+1+1+1	$4 = d_{41}$
	+3+1-1-3	$20 = d_{42}$
	1-1-1+1	$4 = d_{43}$
	+1-3+3-1	$20 = d_{44}$
n = 5	1+1+1+1+1	$5 = d_{51}$
	+2+1+0-1-2	$10 = d_{52}$
	2-1-2-1+2	$14 = d_{53}$
	+1-2+0+2-1	$10 = d_{54}$
	1-4+6-4+1	$70 = d_{55}$

---

* D. B. Delury, Values and Integrals of the Orthogonal Polynomials up to n = 26, published for Ontario Research Foundation by University of Toronto Press (1950).

The procedure then consists of forming a 2nd, 3rd, 4th, ...etc. column of "sums and differences" by applying the appropriate set of linear functions for the number of levels of each successive factor to the preceding column. The final column of "sums and differences" so obtained is squared and divided by an appropriate constant to give the individual degrees of freedom of the analysis of variance.

This individual degrees of freedom come out in the following order:

CF
A_1
A_2
.
.
A_k

B_1
A_1B_1
A_2B_1
.
.
.
A_kB_1

B_2
A_1B_2
A_2B_2
.
.
.
A_kB_2

.
.
.

B_m
A_1B_m
.
.
A_kB_m

C_1
A_1C_1
A_2C_1
.
.
.
.
.
.
.
etc.

for the 4 x 3 x 2 factorial we have

	divisor
CF	24
A_1	120
A_2	24
A_3	120
B_1	16
A_1B_1	80
A_2B_1	16
A_3B_1	80
B_2	48
A_1B_2	240
A_2B_2	48
A_3B_2	240
C	24
A_1C	120
A_2C	24
A_3C	120
B_1C	16
A_1B_1C	80
A_2B_1C	16
A_3B_1C	80
B_2C	48
A_1B_2C	240
A_2B_2C	48
A_3B_2C	240

The column of divisors are those used to divide the corresponding squares of the elements in the 3rd (last column) of "sums and differences".

The appropriate divisors are obtained by writing the coefficients d^{o}_{ij} (given in the table on page 13) in the form

$(d^{o}_{k1}\ d^{o}_{k2}\ \ldots d^{o}_{kk})$ for the k levels of A

and multiplying by the successive terms of

$(d^{o}_{m1}\ d^{o}_{m2}\ \ldots d^{o}_{mm})$ for the m levels of B

to get

$(d^{o}_{k1}\ d^{o}_{m1},\ d^{o}_{k2}\ d^{o}_{m1}\ \ldots d^{o}_{kk}\ d^{o}_{m1},\ d^{o}_{k1}\ d^{o}_{m2} \ldots d^{o}_{kk}\ d^{o}_{mm})$, then

carrying on the process until all factors are accounted for.

For the 4 x 3 x 2 factorial we have:

(4 20 4 20) (3,2,6) (2,2)

Doing the multiplication we get

(12 60 12 60 8 40 8 40 24 120 24 120) (2,2) =

(24 120 24 120 16 80 16 80 48 240 48 240
24 120 24 120 16 80 16 80 48 240 48 240).

as shown on page 15.

Computational check

The sum of squares of the observations is equal to the sum of the individual degrees of freedom, i.e. $\Sigma X^2 = \Sigma \frac{D^2}{d}$.

Combination of individual degrees of freedom

A combination of individual degrees of freedom is accomplished by forming two columns using sets of k (the number of levels of factor A) individual degrees of freedom in the D^2/d column. Two columns are formed: the first by writing the first element of each set of k in sequence the second by adding the next (k-1) elements of each set. The two columns are formed into one by appending the second column to the end of the first. This column is then operated on using sets of m elements (the number of levels of factor B). Two columns are again formed as before and combined. This process is carried out for all factors.

The number of degrees of freedom associated with the final column of "sums of squares" are given by the term of the product

$$(1,k-1)\ (1,m-1)\ (1,r-1)$$

For the 4 x 3 x 2 example we have (from page 15)

CF
B
B
C
BC
BC
A
AB
AB
AC
ABC
ABC

Using sets of 3 from this column we get

CF
C
A
AC
B
BC
AB
ABC

And finally, using sets of 2

CF
A
B
AB
C
AC
BC
ABC

Thus we have

	d.f.
CF	1
A	3
B	2
AB	6
C	1
AC	3
BC	2
ABC	6

The sequence of d.f. is obtained by multiplying

$$(1,3)\ (1,2)\ (1,1) = (1\ 3\ 2\ 6)\ (1,1)$$

$$= (1,3,2,6,1\ 3\ 2\ 6)$$